

# Final Proposed Architectural Design Document

California Minimum Essential Datasets (MEDS)

1/22/2010  
Michael Baker Jr., Inc



Baker

## Contents

1.	Introduction .....	4
1.1	Purpose .....	4
1.2	MEDS Technology Overview .....	5
1.3	Architectural Strategy Statement .....	5
1.4	Abbreviations/Definitions .....	8
1.5	Inputs .....	9
1.6	User Characteristics .....	9
1.7	General Constraints and Risks .....	11
1.8	Assumptions and Dependencies .....	12
2	Recommended Architectural Components .....	13
2.1	Standard Web Mapping Services .....	13
2.1.1	Overview .....	13
2.1.2	Technology Considerations .....	13
2.1.3	Software and Hardware Requirements .....	15
2.2	Customized Web Service Applications (APIs) .....	16
2.2.1	Overview .....	16
2.2.2	Technology Considerations .....	18
2.2.3	Software and Hardware Requirements .....	19
2.3	MEDS Website .....	20
2.3.1	Overview .....	20
2.3.2	Technology Considerations .....	20
2.3.3	Software and Hardware Requirements .....	21
2.4	Alternate Data Delivery .....	22
2.4.1	FTP Sites .....	22
2.4.2	Sneakernet Distribution .....	24
3	MEDS Services Deployment Architecture .....	27
3.1	Overview .....	27
3.2	Technology Considerations .....	27
3.3	Security .....	29

# California Minimum Essential Data Sets Map Services

Document Type: Architectural Design Document

3.4 Implied Requirements.....	29
Requirement Validation .....	29
Figure 1. MEDS Technology Overview .....	6
Figure 2. MEDS System Components.....	7
Figure 3. Simple ESRI ArcServer Architecture (Source: ESRI).....	13
Figure 4. ArcGIS Server Network Load Balancing Configuration Architecture (Source: ESRI) .....	14
Figure 5. Customized Web Service Application Architecture .....	17
Figure 6. One Physical Server Architecture.....	19
Figure 7. Creating Distributed Google Earth Contents for Disconnected Usage .....	25
Figure 8. CA MEDS Deployment Infrastructure .....	28

## **1. Introduction**

### **1.1 Purpose**

This document defines the MEDS Architectural Design based on the MEDS Functional Design Specifications (dated on October 23<sup>rd</sup>, 2009). This Architectural Design document shall not repeat those design specifications. It details the high level key system components and also the deployment architecture to meet the functional design specifications. Key technologies are identified for each MEDS service component.

The MEDS architectural design strives to be both scalable and flexible. The system design is component based and provides great flexibility in deployment.

The technical overview defines major system components and related technologies. It also provides deployment architecture.

The use cases mentioned in this document capture the specific functional and business rules applied on each independent business process activity and should enable the system users to validate the functionalities supported implicitly and explicitly by the system. The use cases also identify the end users of the system.

### **1.2 MEDS Technology Overview**

Driving Baker's architecture design recommendations are the Functional Requirements document which captured California geospatial community input and Steering Committee guidance that access to MEDS data services must be made as easy as possible to the widest possible government audience and that technological barriers to access must be minimized or eliminated. To that end the "technology stacks" that are recommended include substantial open source solutions for maximum accessibility, yet also recognize the large investments in commercial software products and staff training that already exist in government agencies throughout California.

### **1.3 Architectural Strategy Statement**

To meet the MEDS core requirements of a robust, scalable architecture, a combination of technologies has been recommended. We recommend this flexible suite of solutions to take advantage of the strengths of each, while meeting the needs of California's diverse stakeholder community. By providing access to data services and geospatial functionality as open source-based and ESRI-based, we allow MEDS users to consume data and create functionality in whichever way offers the lowest barrier to entry and the easiest implementation in their existing environment. It recognizes the broad appeal of open source to the developer and academic community, while acknowledging the substantial investment with commercial-off-the-shelf GIS software within government agencies across California.

The need to serve MEDS to all technical expertise levels, from end users to developers, further reinforces the strength of a multi-pronged approach to MEDS data services access. Keeping the system compatible with other external data sources via OpenLayers allows visualization of MEDS as an overlay on Google Maps/Earth, Bing or other commonly used systems and gives MEDS the ability to easily mash-up data that may be served by non-MEDS sources.

By wrapping MEDS in an Open Source Web Interface such as Drupal, we allow communication with low-bandwidth protocols, such as RSS, which we see as a potentially significant player in the mobile environment.

# California Minimum Essential Data Sets Map Services

Document Type: Architectural Design Document

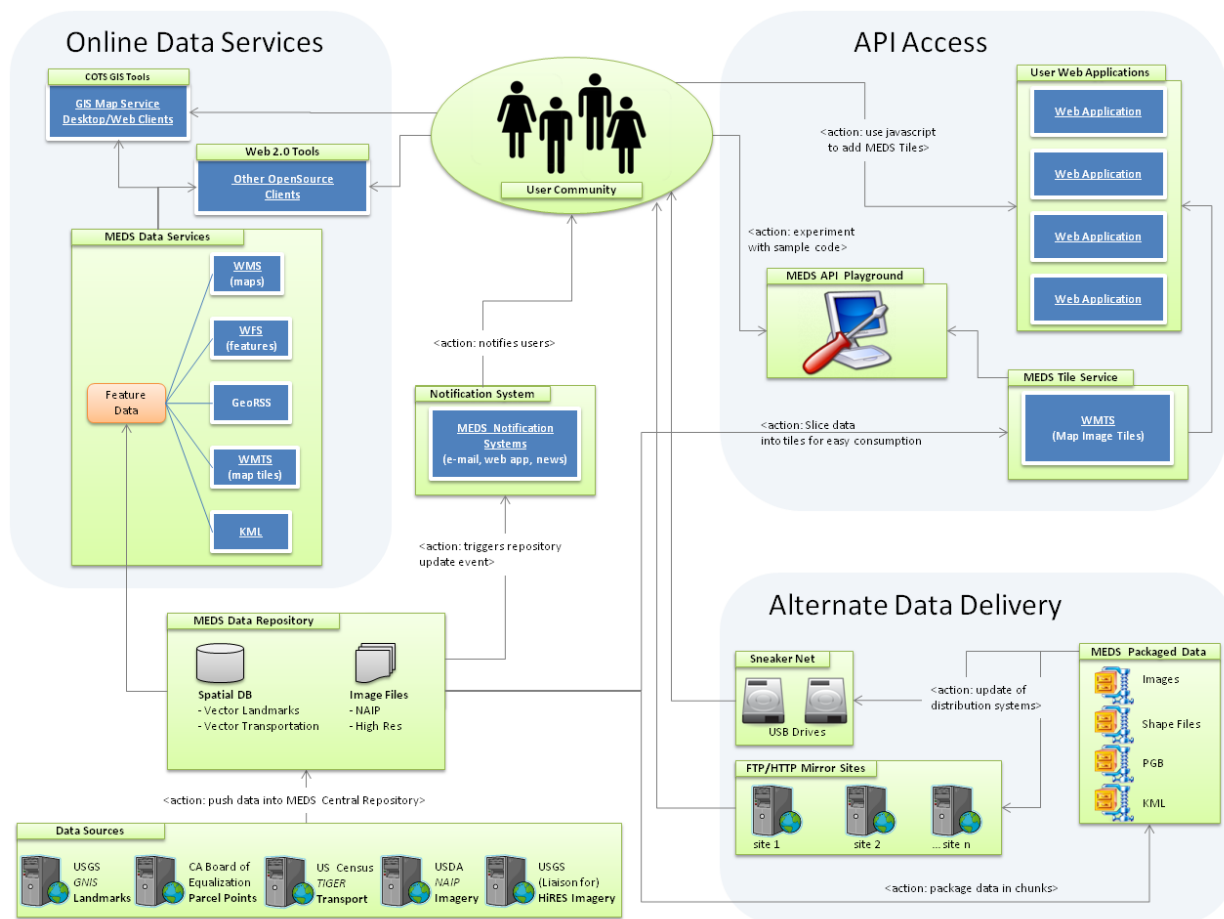


Figure 1. MEDS Technology Overview

This Architecture Design document will describe the technology recommendations to support:

- Hosting the Landmarks, Transport, and Imagery data;
- Enabling the use of commercial-off-the-shelf GIS software for online data services to access the data by the traditional ESRI-centric approach;
- Enabling the development and use of a web-based OpenSource technology stack with task specific APIs to access the online data services;
- Enabling alternate data delivery via FTP for bulk downloads or a Sneakernet when the online data services are down in a catastrophic event.

## California Minimum Essential Data Sets Map Services

Document Type: Architectural Design Document

The primary objectives of the proposed MEDS solutions is to provide a scalable, high performance and maintainable platform for hosting and distributing MEDS data, and also to establish a framework for data providers to contribute to future data improvement and updates. The MEDS service shall include the following key components:

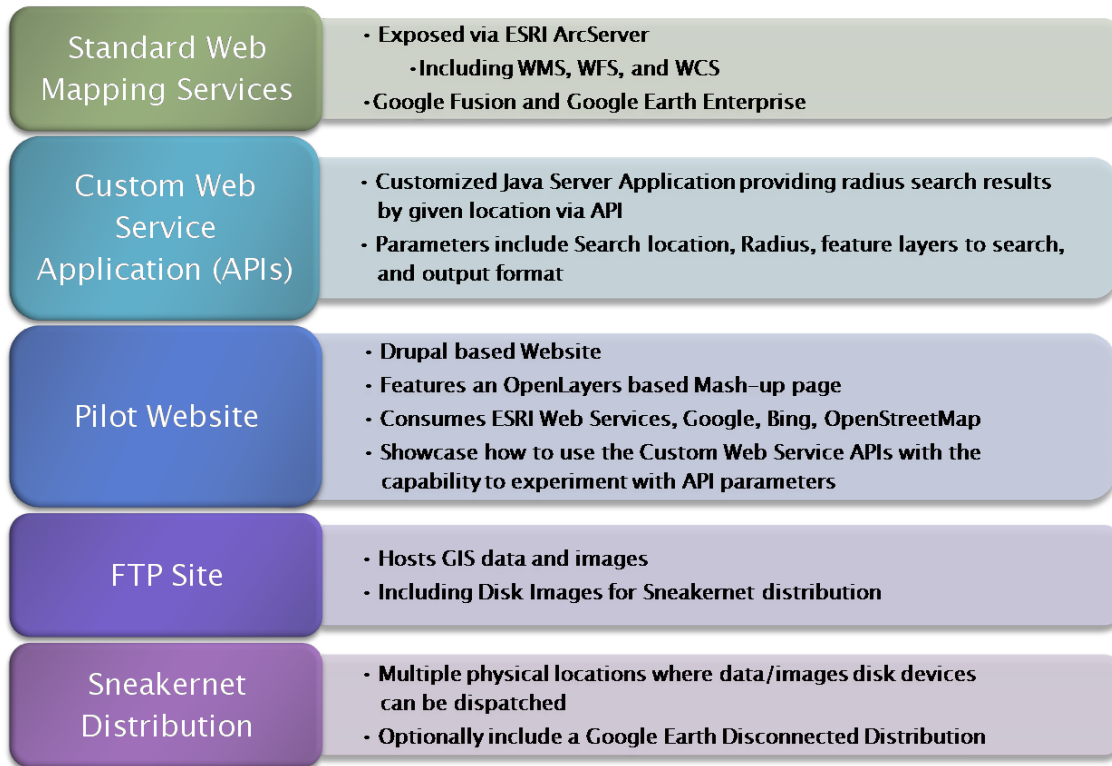


Figure 2. MEDS System Components

## California Minimum Essential Data Sets Map Services

Document Type: Architectural Design Document

### 1.4 Abbreviations/Definitions

Abbreviations and definitions used throughout the document are presented here to aid the reader.

Term	Definition
Actor	An actor in the Unified Modeling Language (UML) "specifies a role played by a user or any other system that interacts with the subject."
AJAX	Shorthand for asynchronous JavaScript and XML is a group of interrelated web development techniques used on the client-side to create interactive web applications or rich Internet applications.
API	Application Programming Interface
COTS	Commercial Off-The-Shelf
Cloud Computing	A paradigm of computing in which dynamically scalable and often virtualized resources are provided as a service over the Internet
Crowd Sourcing	A neologism for the act of taking tasks traditionally performed by an employee or contractor, and outsourcing it to an undefined, generally large group of people or community in the form of an open call
CSW	Web Catalog Service: Access to catalog information
DHS	Department of Homeland Security
EOC	Emergency Operations Centers
EROS Center	USGS Earth Resources Observation and Science Center
ETL	Extract, Transform, and Load
FME	A GIS ETL tool by SAFE Software
FTP	File Transfer Protocol
GEORSS	An emerging standard for encoding location as part of a Web feed.
GeoXACML	Geospatial eXtensible Access Control Markup Language (currently being standardized)
GML	Geography Markup Language: XML format for geographical information.
GUI	Graphical User Interface
HDDS	USGS Hazards Data Distribution System
HTTP	Hyper Text Transfer Protocol
KML	Keyhole Markup Language
MEDS	Minimum Essential Datasets
NLB	Network Load Balancing
OGC	Open Geospatial Consortium
RSS	Really Simple Syndication
SDSC	San Diego Super Computer Center
SFS	Simple Features – SQL
SLA	Service Level Agreement
SOM	ESRI Service Object Manager
SOC	ESRI Service Objects Container
SRS	Software Requirement Specification
SMS	Short Message Service
UML	Unified Modeling Language
WCS	Web Coverage Service: Provide coverage objects from a specified region.

## California Minimum Essential Data Sets Map Services

Document Type: Architectural Design Document

Web 2.0	Commonly associated with web development and web design that facilitates interactive information sharing, interoperability, user-centered design and collaboration on the World Wide Web.
WFS	Web Feature Services
WFT	Web Feature Transactions
WMS	Web Mapping Services
WMTS	Web Mapping Tile Services
WPS	Web Processing Service: Remote processing service

### 1.5 Inputs

MEDS Final Functional Requirements (dated October 2<sup>nd</sup>, 2009).

MEDS Final Design Specifications (dated October 23<sup>rd</sup>, 2009).

MEDS stakeholder meeting notes.

MEDS 09/14/09 Steering Committee meeting notes.

### 1.6 User Characteristics

The use cases mentioned below capture the specific functional and business rules applied on each independent business process activity and should enable the system users to validate the functionalities supported implicitly and explicitly by the system. The use cases also identify the end users of the system.

The initial users of MEDS data and map services are government GIS professionals or technical staff who should have a working knowledge of GIS and spatial databases. MEDS initial users are emergency response personnel (Emergency Operations Centers (EOC)). Since MEDS architecture also supports spatially enabled RDBMS and customized Web Service APIs, it is anticipated that more non-traditional GIS programmers would extend the MEDS's reach to an even broader audience.

Actor	Description
Metropolitan EOCs (User)	<p>It is anticipated that urban, well-funded EOCs with adequate internal IT infrastructure will prefer to rely on data stored within their own networks in order to minimize disruption in case of external connectivity failure. Primary benefits of MEDS for this type of client will be as a backup repository for their own local data to be used for disaster recovery and as an easily accessed source of data for adjacent jurisdictions.</p> <p>Normal Use</p> <p>Daily use would consist of checking MEDS for updates and</p>

## California Minimum Essential Data Sets Map Services

Document Type: Architectural Design Document

	<p>downloading updated datasets for adjacent jurisdictions. Uploading new data from local sources to MEDS would be considered routine maintenance. Normal use might also include running HAZUS and other models, developing evacuation plans and other pre-planning activities.</p> <p>Event Use</p> <p>Assuming that the EOC has up-to-date data when an event occurs, their event-driven usage of MEDS may emphasize access to post-event data. Download of or service-based access to post-event imagery could be a critical use. Imagery could be loaded once and used by many agencies. Pushing high-priority event-based local data up to MEDS in order to share the common operating picture (COP) with other local agencies, adjacent jurisdictions and state agencies could be beneficial.</p>
Rural EOCs (User)	<p>EOCs that lack internal capacity for data storage or processing may see benefit in using MEDS services more directly for data storage, access and applications.</p> <p>Normal Use</p> <p>Daily use would consist of accessing the MEDS data sets for disaster planning activities. Uploading new data from local sources to MEDS would be considered routine maintenance.</p> <p>Event Use</p> <p>Adequate bandwidth is a concern for some rural jurisdictions. A MEDS “data pack” consisting of hard drives with relevant MEDS data that can be rapidly deployed to responding agencies could help mitigate this concern.</p> <p>All EOCs will have a need to create maps for distribution to field crews, public officials, and the press. MEDS data layers can serve as a common, easily accessed backdrop.</p>
CalEMA State Offices (User)	<p>One primary benefit of MEDS to state-level agencies will be the pervasive availability of easily accessed data sets. This should make the construction of common operating pictures significantly easier across organizational boundaries.</p>

## California Minimum Essential Data Sets Map Services

Document Type: Architectural Design Document

	<p>Normal Use</p> <p>MEDS data will be used for disaster planning activities.</p> <p>Event Use</p> <p>During and immediately after an event, CalEMA staff can use MEDS data in combination with other datasets to assist with disaster assessment and recovery. It is anticipated that CalEMA staff will use the web-based data services to access MEDS, rather than needing a local copy. CalEMA staff may play a role in pushing post-event data (event-related landmarks, for instance) into MEDS.</p>
State/Federal Agencies (User)	<p>Normal Use</p> <p>Routine pushes from data stewards (e.g., Caltrans, USGS) will be required to keep some data sets updated. Development of relevant functionality, such as buffer analysis of landmarks (e.g., query all schools within 1 mile of an event location) could lead to routine usage of MEDS data. This would be true for county and regional level normal use as well.</p> <p>Event Use</p> <p>Post-event data pushes will be required, especially for imagery. State, regional, and federal agencies may be able to use MEDS landmarks to enhance their COPs.</p>
Mash-Up Developers (Customized API Users)	<p>Software developers who consume Web Services (including MEDS Web Service APIs) to create mash-up applications.</p>
Mash-Up Application Users (users of mash-up applications)	<p>End users who benefits from MEDS through the mash-up applications that consume MEDS Web Services.</p>

### 1.7 General Constraints and Risks

1. MEDS Web Services require constant and reliable internet connectivity and some data services demand broad bandwidth. Some rural areas do not have reliable internet connectivity and bandwidth is limited. Any system design shall consider this constraint and provide alternatives.

2. The majority of data stakeholders and professional GIS users of the targeted government agencies use ESRI technologies. Tremendous investment has been made in these technologies in creating and maintaining spatial data, providing data analysis and visualization functions, and in offering Web Mapping Services (WMS, WFS, and WCS); and some of those mapping services utilize domain specific geoprocessors. Most staff's GIS training has also been ESRI-centric. It is important to consider this constraint factor in architectural and application designs.
3. MEDS map services have a high level of up-time requirement. They are required to be 24x7, even though no specific SLA (Service Level Agreement) up-time is defined. Cost is an obvious constraint factor. Any data storage and web service design needs to consider the high SLA requirement.

### **1.8 Assumptions and Dependencies**

- 1 It is assumed that two core data centers will be available for data storage and web servicing hosting: NASA Ames Research Center and San Diego Supercomputer Center.
- 2 It is assumed that a number of mirrored sites would be available for MEDS data downloads: Cal-Atlas may be an appropriate mirror site for downloads. Other potential mirror sites might include the Hazard Data Distribution System (HDDS), NIFC, Geomac, and others.
- 3 It is assumed that all public web services such as Google Maps, Bing Maps, and Open Street Maps shall be available. MEDS solutions have no control over those services in terms of SLA (Service Level Agreement). There is an inherent risk of using any public service.
- 4 It is assumed that there is a steward for each of the MEDS dataset. Per the Steering Committee Meeting Notes, CalEMA will be the Landmarks steward, Caltrans will be the Transportation steward and the USGS will be the Imagery steward.
- 5 While the data stewards might also serve as hosts for their MEDS data, this is not a required role. A more robust solution may be for the stewards to simply maintain data and to regularly push their data to the core data centers for distribution. Benefits of this approach include taking advantage of the up-time, performance and bandwidth of the core data centers, reducing load on the stewards' infrastructure, and simplifying security and administration of the MEDS.

## 2 Recommended Architectural Components

### 2.1 Standard Web Mapping Services

#### 2.1.1 Overview

MEDS includes both vector and raster datasets. Those data shall be exposed as OGC-compliant Web Mapping Services such as WMS, WFS, and WCS. The targeted users for those services are the professional GIS and mapping professionals who have either client or server tools to consume them. Those users can utilize the built-in functions of their existing commercial off-the-shelf GIS tools to interact with the MEDS Web Mapping Services along with other datasets.

There are two key components to those Mapping Services. One is the enterprise strength and spatially enabled relationship base to store, version, and archive both spatial and attribute database, and the other is an expandable server application framework that enables rapid publications of the Mapping Services.

#### 2.1.2 Technology Considerations

Based on the MEDS Functional Design Specifications document, MEDS shall build and publish the standard Web Mapping Services on the ESRI ArcServer stack. This includes both the spatial data management component (Arc SDE) and the application server framework component (ArcServer). The ESRI ArcServer framework can be expanded to support domain specific geoprocessors. Initially MEDS can have a simple architecture (Figure 3).

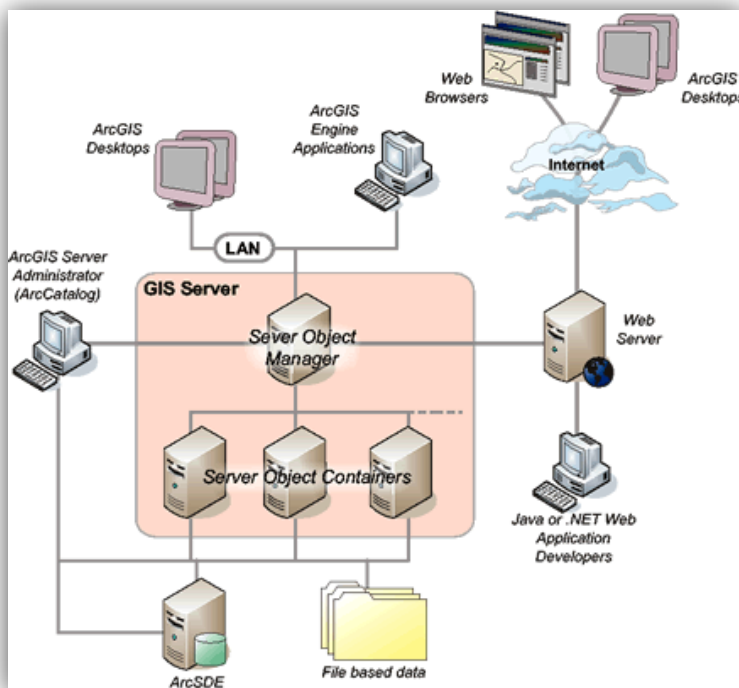


Figure 3. Simple ESRI ArcServer Architecture (Source: ESRI)

## California Minimum Essential Data Sets Map Services

Document Type: Architectural Design Document

When user demands increase MEDS shall expand to provide more robust services using a Network Load Balancing Configurations (Figure 4).

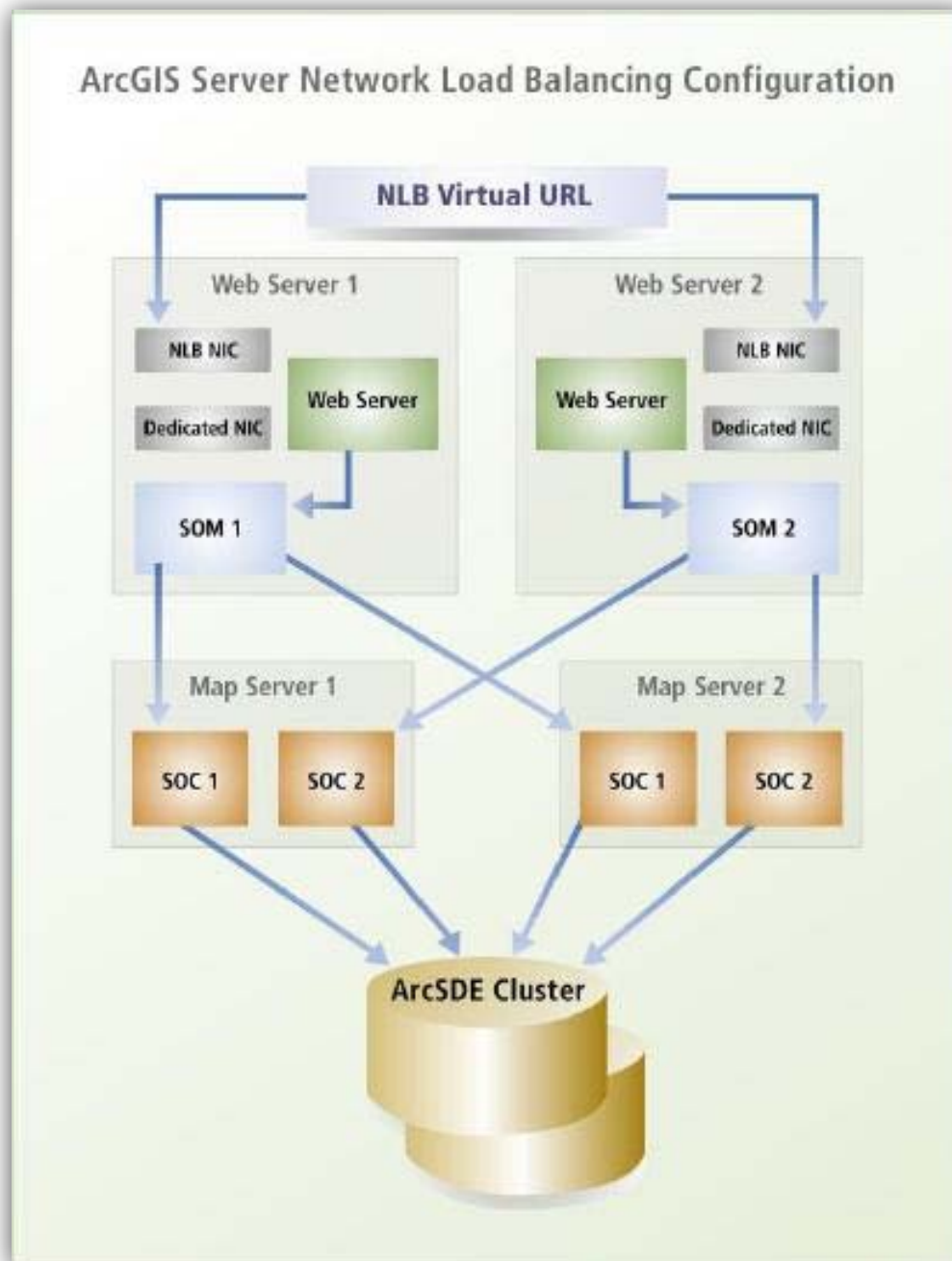


Figure 4. ArcGIS Server Network Load Balancing Configuration Architecture (Source: ESRI)

### **2.1.3 Software and Hardware Requirements**

The following requirements shall form the starting point for building a basic ArcServer architecture. At the minimum One Single Server can include all system components. A Windows slant is chosen since it is used by most professional IT staff (hence a bigger pool of supporting staff). There is no technical reason not to choose a UNIX implementation.

- 1) ArcServer 9.3 for Windows.
- 2) Oracle 10g or 11g, or Microsoft SQL 2005 or above.
- 3) Microsoft IIS 6 or above.
- 4) A quad core (Intel or AMD) server with minimum 8 GB RAM.
- 5) Minimum 1 TB database storage space.
- 6) Minimum 2 TB for file storage space.

## **2.2 Customized Web Service Applications (APIs)**

### **2.2.1 Overview**

While professional GIS and mapping users can quickly utilize the Web Mapping Services published via ArcServer, many non-GIS programmers find Web 2.0 type of Web Service much easier to understand and implement, especially in conjunction with other publically available services such as Google Maps, Bing Maps, Yahoo Maps, and Open Street Maps. Those programmers, and by extension, the users of their applications, desire a much simple and easier access to spatially-enabled database.

This shift or expansion of GIS and Mapping services into Web 2.0 technologies also applies to the Web Mapping Services. Web Mapping Tile Service (WMTS) is being adopted much more broadly and at a faster pace than the traditional WMS and WFS, and WCS.

Even though early MEDS users will be GIS professionals at government agencies, who probably would use WMS for imagery in their commercial off-the-shelf (COTS) GIS client tools, there is also a requirement for all types of users to be able to visualize MEDS images on popular mapping engines such as Google Maps and Bing Maps. MEDS will provide such a service. Users may find that accessing MEDS in this way can result in faster display time, particularly where targeted functionality is desired.

A set of tiles of the MEDS images could be created, hosted on the data centers, and made available via customized Web Service Applications (APIs). Users do not need professional client tools to overlay those cached tiles on Google Maps or Bing Maps.

Those cached tiles shall also be included in the Sneakernet distribution. Optionally this set of tiles can also be offered to commercial hosting services such as Google and Microsoft.

There are also some domain specific functions that non-GIS programmers may desire. For example, a Web Service can be developed to directly interface with spatially enabled database, bypassing the traditional management layer for better performance and simplicity. Domain-specific functions to take advantage of this type of flexible and fast access to geospatial databases will need to be developed by the early adopters of MEDS. MEDS shall establish an infrastructure on such functions can be built and expanded. Ideally the OCIO Office can work with data stewards to identify what functions shall be built first to have some early successes.

## California Minimum Essential Data Sets Map Services

Document Type: Architectural Design Document

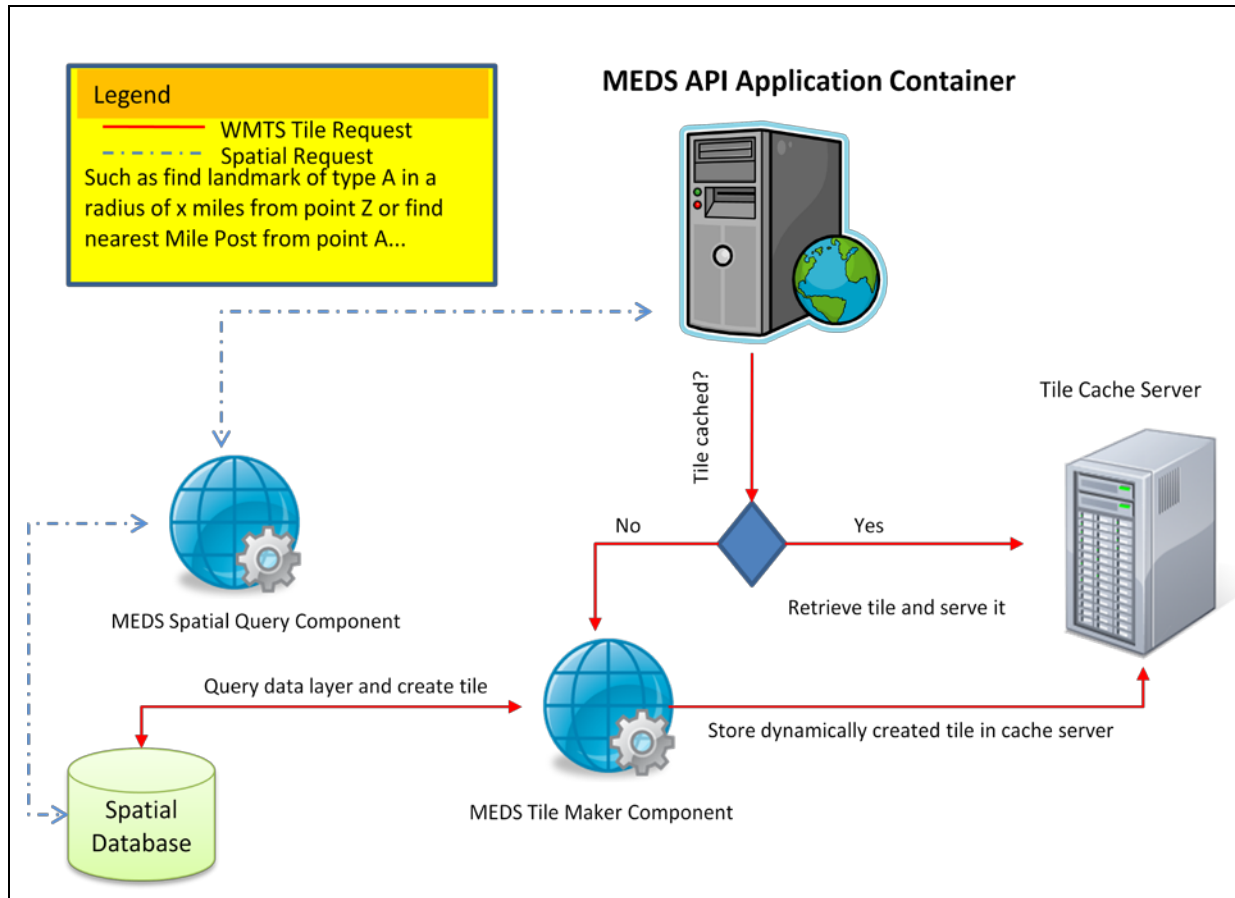


Figure 5. Customized Web Service Application Architecture

The architecture (Figure 5 above) features a Web Service Application Container, which has two key components. First component is WMTS Tile Maker, which serves and/or creates tiles of feature layers from spatially-enabled RDBMS. For example, when a mash-up developer requests tiles of a specific feature layer such as landmarks, this component will check if the requested tiles are already pre-cached and sent back with valid URLs. If not, this component will dynamically create tiles from the spatially-enabled database, save them in the cache server and send the valid URLs back to the requester.

The second component is spatial query component that provide either SOAP and/or REST API calls to perform spatial queries on the spatially-enabled RDBMS, and send results back to requesters. For example, a mash-up developer can send this component a location and ask for a list of schools within 5 miles, this spatial query component would do the query in the database and return the record set to the requester in XML/JSON stream. Multiple spatial query components can be built to provide domain specific functions.

The sample code for using the cached tiles and domain specific functions shall also be made available in a “Code Playground” type of environment. The sample code shall demonstrate how MEDS tiles and functions can be used in all major online mapping engines such as Google Maps,

Bing Maps, and Yahoo Maps. Special efforts shall be made to demonstrate the uses of MEDS tiles on Open Source engines such as OpenLayers.

### **2.2.2 Technology Considerations**

There are two distinguished technology components: The spatially-enabled database with a proven track record in stability, scalability, and performance; and an application development framework on which more Web Services can be built upon.

The spatially enabled database shall be an independent relationship database and shall have the data synchronized with the ArcSDE as described in Standard Web Mapping Services section. This independency may appear to be redundant but it reduces the potential interference with the ArcServer implementation. It also contributes to additional MEDS system reliability.

There are at least two options to choose from as the spatially enabled RDBMS. One is Oracle Spatial, which has a proven track record in large enterprises and agencies, and PostgreSQL with PostGIS, which is Open Sourced and popular in the Web 2.0 development community. MEDS may want to start out with Post GIS and evaluate the Oracle alternative when user demands rise and/or shift.

A multiple application framework is also possible and could include players such as Microsoft .NET, PHP, and Java. Java works well with both Oracle and PostgreSQL, and also has more options in application hosting services, so we would recommend Java as a logical place to build the foundation of the application framework.

Figure 6 features a simple one physical server architecture, even though it can be easily expanded to support Network Load Balancing. Ideally there shall be at least two data centers hosting both data and applications for fail-safe purposes and for added scalability and performance.

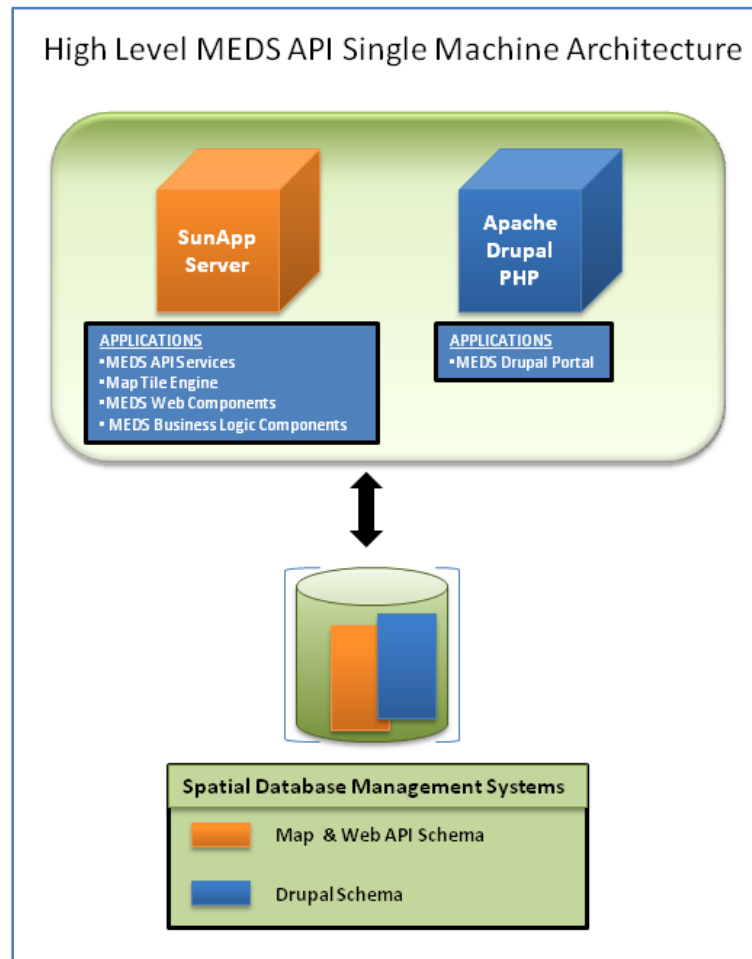


Figure 6. Single Server Architecture

There shall be a process to notify the consumers of the web services in unlikely event of a catastrophic failure of either or both data centers. A monitoring application shall be developed and hosted outside data centers to check on the system down-time and send emails, RSS feeds and optionally SMSs to both technical support staff and web service users.

MEDS shall also provide an API Playground, modeled after Google API and Microsoft API playgrounds. Users shall be able to edit sample code interfacing with the APIs and see how they work. Users shall also be able to save and copy/paste the source code.

### 2.2.3 Software and Hardware Requirements

The following requirements shall form the starting point for building a basic Web Service Application Server architecture. At the minimum One Single Server can include all system components.

- 1) Sun Application Server 9.x or above for Windows.
- 2) Glassfish Web Server 3.x or above for Windows.

- 3) PostgreSQL 8.4 or above with Post GIS for Windows.
- 4) A quad core (Intel or AMD) server with minimum 8 GB RAM.
- 5) Minimum 1 TB database storage space.
- 6) Minimum 1 TB file storage space.

## **2.3 MEDS Website**

### **2.3.1 Overview**

There shall be a dedicated MEDS website to document and promote MEDS data and Web Services. It shall also form a base for an active user community. At minimum there shall have the following components.

- 1) Data documentation.
- 2) Data discovery (with map interface).
- 3) Data distributions (links to file downloads).
- 4) Web Service API Playground (including sample codes and documentation).
- 5) News and Project Status (RSS and GEORSS).
- 6) Organizational information (MEDS Steering Committee, Key data stewards, etc.).
- 7) User Community (Calendars, Polls, Discussion Threads, Blogs, Tweets...).

The MEDS website shall start as an actively supported site with strong focus on data stewards and data users. As both MEDS data and Web Services grow, increasing attention shall be given to the non-GIS programmers. Traditional GIS programmers and professional GIS and mapping users should be able to understand the MEDS data and services with less effort, and may be more interested in data updates.

### **2.3.2 Technology Considerations**

There are many well supported and robust Content Management tools that are capable of powering a MEDS website. The selection of a particular tool is largely based on the hosting services and the ease of maintenance. Drupal is such a popular and user friendly Open Sourced tool and is currently used at AMES. Drupal has many built-in modules to provide the functions as required by the above components such as RSS aggregator, calendars, polls, discussion threads, and blogs.

ArcServer can expose the data catalog service that is needed for the data documentation and data discovery. Drupal can provide a frame to wrap such a module.

### **2.3.3 Software and Hardware Requirements**

Drupal has broad supports for Operating Systems, backend RDBMS, and web servers. The following list is based on single physical server architecture in Windows environment.

- 1) Drupal 5.2 or above for Windows.
- 2) Apache 2.2 or above for Windows.
- 3) Python 2.5 or above for Windows.
- 4) PostgreSQL 8.4 or above with Post GIS for Windows.
- 5) PHP 5 or above for Windows.
- 6) A quad core (Intel or AMD) server with minimum 8 GB RAM.
- 7) Minimum 500 GB database storage space.
- 8) Minimum 1 TB file storage space.

## **2.4 Alternate Data Delivery**

### **2.4.1 FTP Sites**

#### **2.4.1.1 Overview**

MEDS includes primarily spatial datasets, both vector and raster. All MEDS vector data are provided as ESRI formats or one of the formats that can be translated into an ESRI format by the data owners. While most of images (aerial and/or satellite photos) initially are provided as files, pyramided files or tiles can be created to enable faster loading in most mainstream GIS and mapping tools.

Vector data storage shall be in an enterprise spatial database with versioning capability regardless of how the source data is provided. This spatial database shall serve as the official MEDS database for the vector data. A set of GIS files such as shape file, Personal GeoDatabases, or KML files shall be created and zipped by county level from this spatial database.

Images shall be stored at the highest resolution and lowest possible compression as geo-referenced image files. There shall be a mechanism to record the metadata of each image. A set of pyramided image files can be created for a smaller geography, and a set of lower resolution of images can be created for quick file loading.

A set of pre-defined database and file packages shall be developed to reduce both distribution confusion and distribution size. For the vector data, files shall be created at the county level and can be aggregated to create regional datasets. For the image data, files shall also be organized at the county level

The official storage of data and files shall be on at least two data centers for fail-safe purpose.

#### **2.4.1.2 Technology Considerations**

FTP is the preferred method for large dataset and file download due to its better transfer reliability over HTTP download method. For smaller dataset and files HTTP shall also be supported.

In addition to a minimum of two core data centers, there shall be multiple mirrored FTP and HTTP download sites in California to increase the download reliability and speed. There shall be at least three additional FTP and HTTP mirrored sites in the state on reputable data centers such as those operated by large government agencies and universities. The OCIO shall encourage additional mirrored service while maintaining a list of endorsed sites.

The mirrored sites shall also consider the download needs from Central and Eastern US (for Federal access). MEDS shall explore the possibility of leveraging existing governmental data hosting infrastructure such as DHS and Eros Data Center via HDDS.

There is no additional security requirement for file downloading. Required login is designed to restrict file access in order to limit downloading to government users to control bandwidth requirements to serve the MEDS stakeholders.

MEDS data discovery shall be made easy via proper channels such as Web page presence, RSS and GEORSS feeds, and email list. Most popular geospatial data aggregators such as Geospatial One Stop and HDDS can consume and publish standard RSS and GEORSS feeds.

### ***2.4.1.3 Software and Hardware Requirements***

The selection of FTP server software is largely decided by the hosting services. The only requirement is that it enables secured login. There is no preferred FTP server software recommended. All FTP services shall be download-only (One Way). MEDS shall not provide any FTP client software. Users are free to use FTP client software of their choice. The following lists the minimum requirements for any MEDS Data and File FTP Sites:

- 1) Minimum 1 TB storage space.
- 2) Minimum T1 upstream speed.
- 3) Any FTP server software that allows secured FTP access.
- 4) Microsoft IIS FTP acceptable.

### **2.4.2 Sneakernet Distribution**

#### **2.4.2.1 Overview**

Sneakernet distribution of MEDS data is crucial to the user communities that have either no or low internet connectivity. This applies to both geography areas that have poor internet services or in the situation that network is either overloaded or nonexistent. The latter is called Disconnected (or un-tethered) Mode.

In order for the Sneakernet Distribution to be useful and effective, the following must be considered:

The data must be in sync with data available via the Standard Web Mapping Services and Customized Web Application Services.

- 1 Data must be ready to go.
- 2 Users can reach to distribution centers quickly.
- 3 Users can use data in their own environment immediately once the data arrives.

#### **3.1.1.1 Technology Considerations**

There are two methods by which MEDS can provide the Sneakernet Distribution, and both shall be delivered on high speed external storage devices (USB 2.0 and above).

One option is to create a data file package that includes all PGDB files, image files, and also a MXD file that lists all feature layers. A GIS user can connect the device to a host machine and load the data. This option assumes that an end user has ESRI ArcMap, ArcReader, or ArcExplorer.

Another option is to create a distributed Google Earth deployment for disconnected usage. Essentially all MEDS data (vector and raster) can be loaded into Google Earth Server software. VMWare is then used to create a Virtual Machine with Google Earth software (client software and server software plus loaded data). A VMWare Player and the virtual machine image can be copied to a USB drive and distributed to an end user. The device can then be connected to a host PC, and user can launch the VMWare Player, which launches the virtual machine.

It should be recognized that significant time may be required to prepare and process data in Google Fusion and to publish the data. While the processing time poses minor issues for data sets which can be routinely processed ahead of time, turn-around times on event-based imagery will need to be carefully considered and managed in order to provide data in a timely fashion.

This option sounds complicated, but actually once the USB drive is delivered to an end user, he or she can have the latest MEDS data within the Google Earth environment and all Google Earth functions are available.

## California Minimum Essential Data Sets Map Services

Document Type: Architectural Design Document

The following diagram (Figure 7) illustrates the process of building the distributed Google Earth contents for disconnected usage.

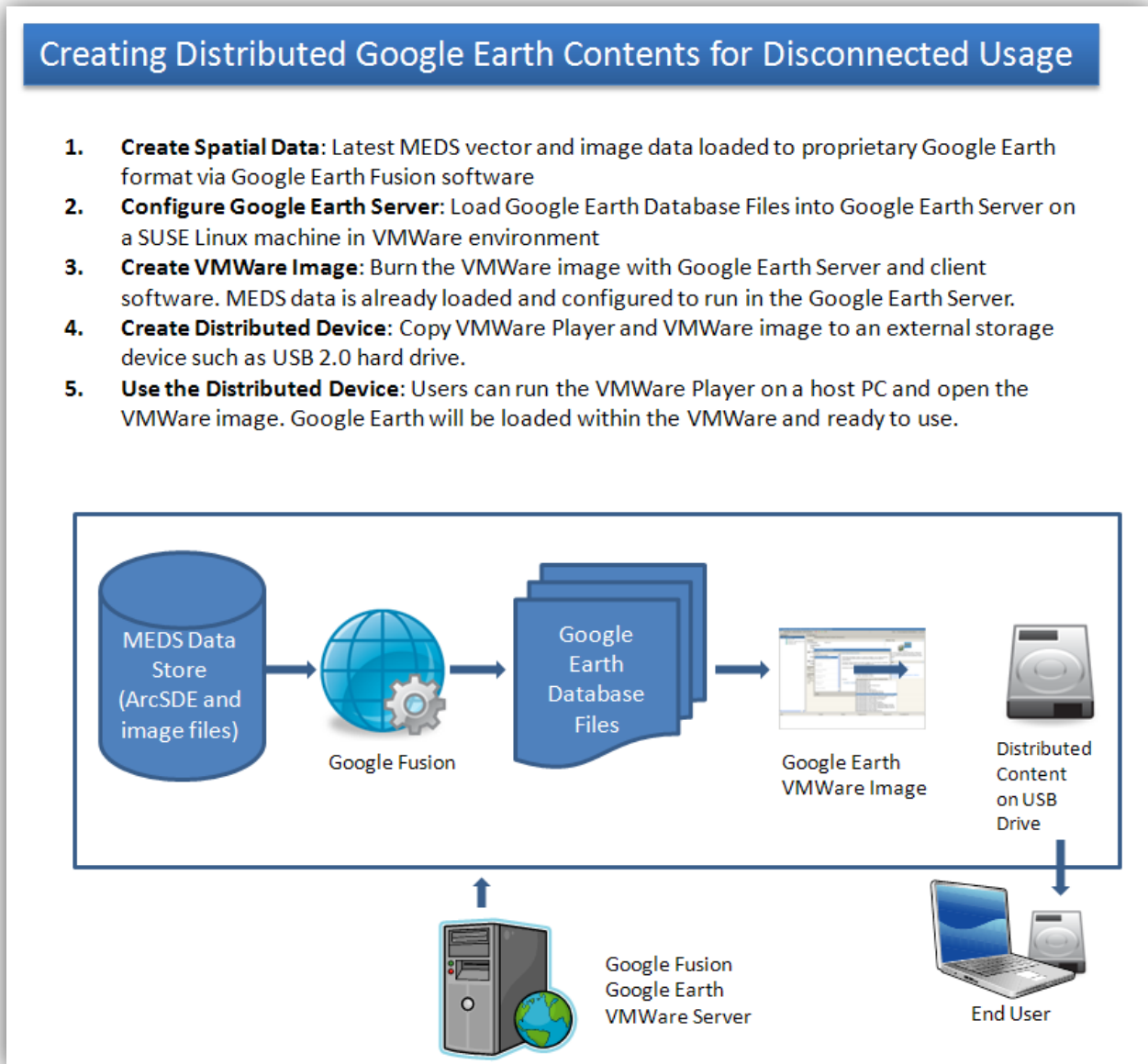


Figure 7. Creating Distributed Google Earth Contents for Disconnected Usage

Three data distribution centers shall be created in Northern, Central, and Southern California to reduce delivery time during emergency. A complete set of data shall also be available in the Central and Eastern US.

### 3.1.1.2 Software and Hardware Requirements

For the data and image file distribution option there is no specific software requirement. All data and image files (and other supporting files such as release note) are copied into a directory

on a production machine once there is a data update. After that all content in that directory can be copied to a number of external hard drives. Each drive should have 1.5 TB or more storage space.

For Disconnected Google Earth deployment option, it is assumed that a valid Google Earth Enterprise license is available for an end user. The following lists the requirements for single physical server architecture to create needed Google Earth distribution content:

- 1 Minimum 2 Intel 3.0 GHz or 2 AMD Opteron 248 CPUs.
- 2 Minimum 1 GB RAM per core; 2 GB RAM per core recommended.
- 3 1 Ethernet Network Interface Controller (Gigabit Ethernet recommended).
- 4 Supported Linux operating systems:
  - SuSE Linux Enterprise Server version 9 and 10 (32-bit and 64-bit)
  - Red Hat Enterprise Linux, AS and ES, version 4 (32-bit and 64-bit)
- 5 Minimum 1 TB storage requirements.
- 6 VMWare Server 2.0 or above.
- 7 VMWare Player 3.0 or above.

### **3 MEDS Services Deployment Architecture**

#### **3.1 Overview**

MEDS architecture is component and module based, which makes the deployment more flexible. It is possible for multiple hosting services to “pick and choose” which components and modules to host. One hosting service may elect to host multiple components such as FTP services and MEDS Website, while a component such as FTP services can be hosted by multiple service providers for added reliability.

This deployment flexibility cannot come at the cost of reliability of interactions among components. It is recommended that MEDS solutions shall start in a more controlled environment and focus more on system reliability. As needs rise and also more confidence gained about the reliable interactions among different hosting services, the deployment architecture can expand.

#### **3.2 Technology Considerations.**

At minimum, all MEDS components can reside on single physical server. While this set up is economical and fast to deploy, the system performance and reliability may be at risk.

Overall, for each key MEDS component there shall be redundancy via Load Balancing or mirror capability (FTP sites for examples). It is more important to have multiple hosting services for the same components than having one or two data centers to host every component. From user’s perspective, the MEDS Website is the point of entry to understand the latest deployment architecture. For example, users can find the updated list of FTP sites (including mirrors). For the mash-up developers the Web Service calls shall be transparent and application loading balancing shall automatically route the calls to appropriate server.

The following diagram illustrates the flexible deployment architecture.

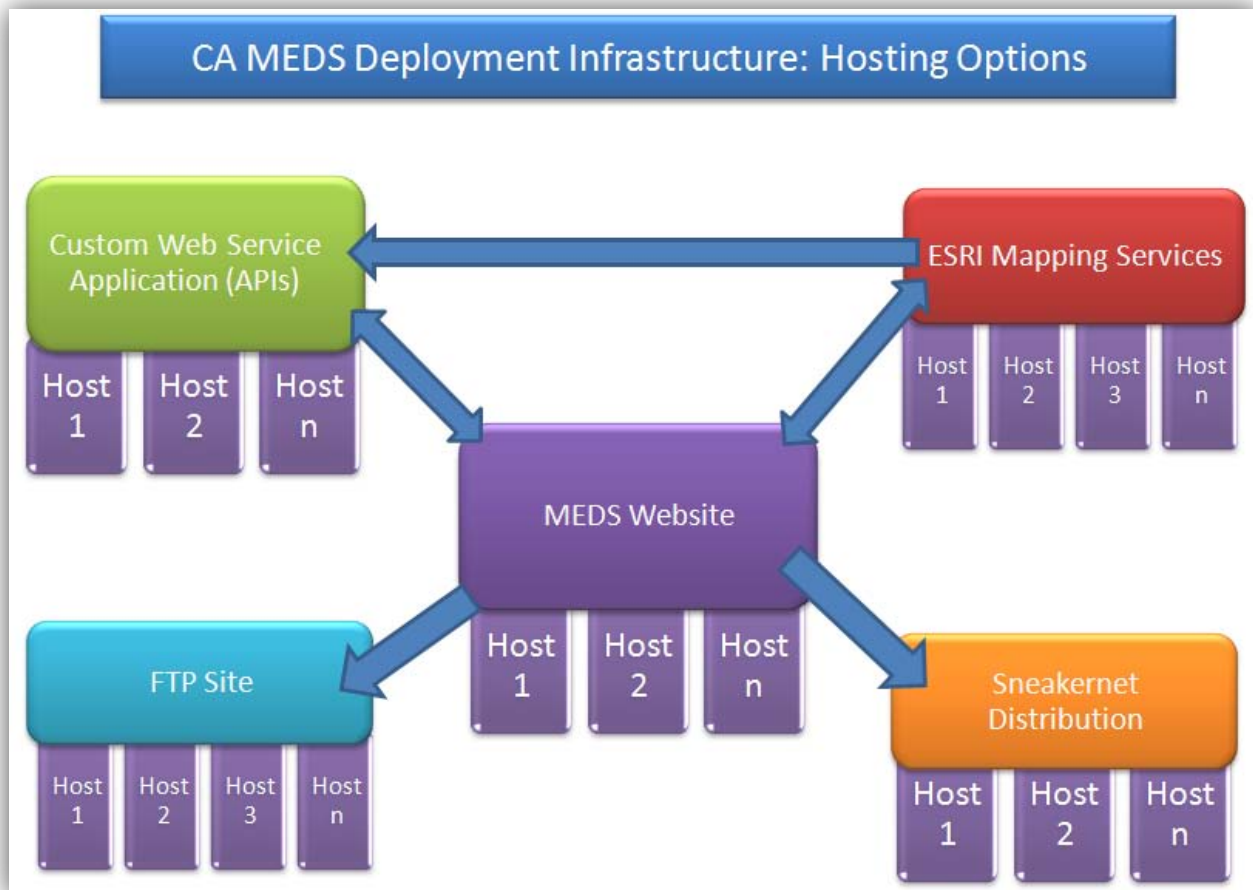


Figure 8. CA MEDS Deployment Infrastructure

The middle ground could be that MEDS starts with a three server set up:

- 1 Web Server: Hosting MEDS website and FTP site.
- 2 ESRI Stack Server: Hosting both data and ESRI Mapping Services
- 3 Customized Web Service Application Server: Hosting customized applications and a spatially-enabled RDBMS.

Performance measurements can be conducted on each server to determine which one needs to be scale up with additional servers.

### **3.3 Security**

Users of MEDS will be government agencies, and both data and mapping services shall be provided in a secured way. All MEDS database and application activities shall be in compliance with hosting service security requirements.

From a data and mapping services perspective, MEDS shall not require an additional security layer and can exist in the public web server layer and shall be read-only. In case the GIO Office decides to allow multiple hosting services for different components, no additional security layer shall be required other than meeting local security requirements.

From data and API access perspective, a simple security measure shall be implemented that checks user name, password, and domain name/IP address. No additional security shall be needed since all data in MEDS can be obtained publically. The security measure is provided with the objective of restricting access in order to provide better performance for the selected group.

### **3.4 Implied Requirements**

- 1 User friendly GUI for any end –user applications.
- 2 OGC compliance in Mapping Services.
- 3 Metadata available for all MEDS datasets and in compliance with FGDC.
- 4 All users shall have FTP software when and if they access to the MEDS data via FTP software.
- 5 Users who desire the disconnected Google Earth deployment shall have Google Earth Enterprise license in place.

### **Requirement Validation**

This architectural design document is based on the functional design specification document dated October 23<sup>rd</sup>, 2009. Changes in the requirements would have impacts on this document. Both documents shall be updated and stay in sync.